What Do Short Bandwidth Probes Tell Us?

Rich Wolski

Martin Swany

UC Santa Barbara

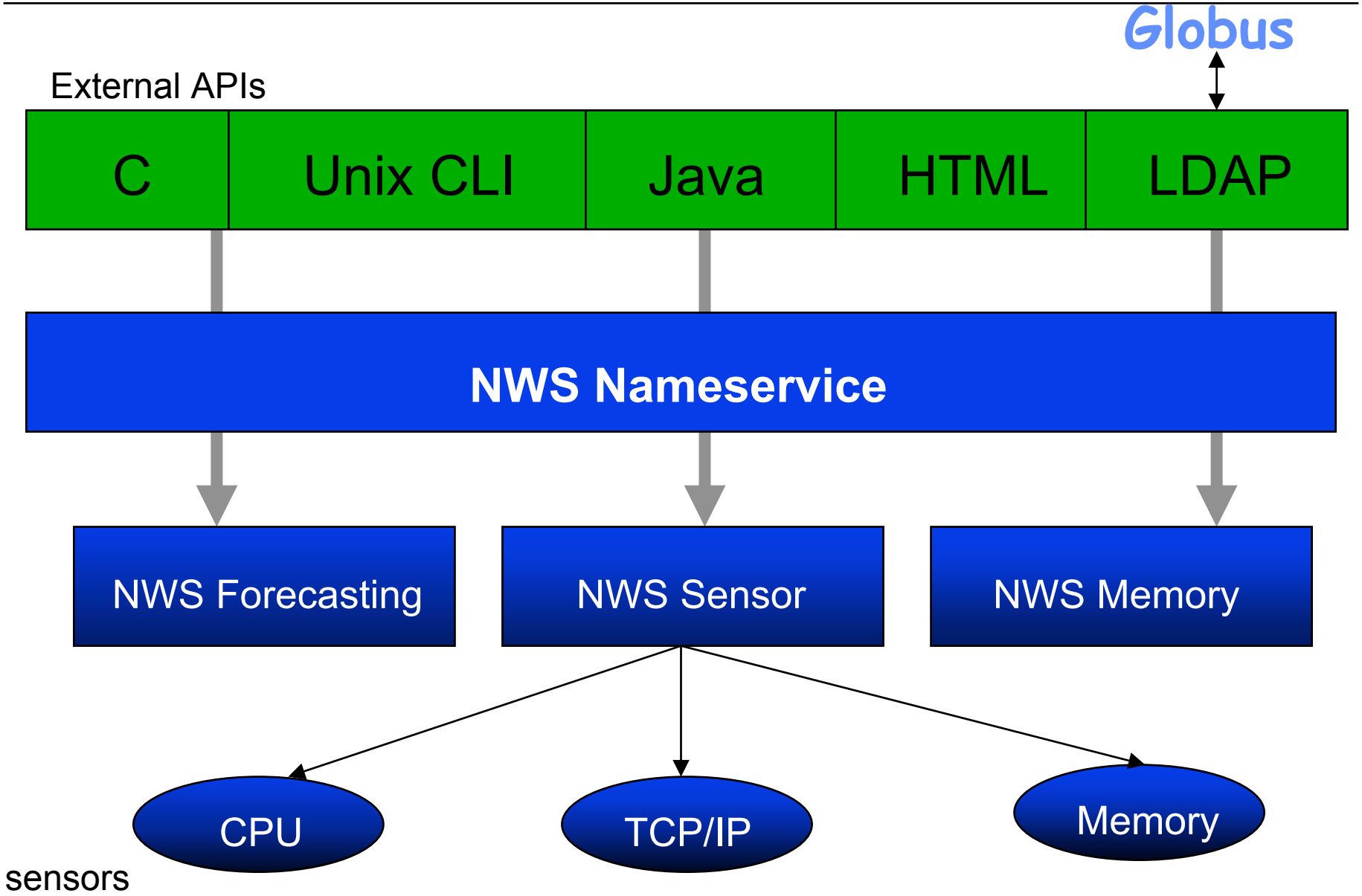
Network Performance Prediction

- Use statistical analysis of previously observed network performance data to
  - Derive distributions used in random simulations
  - Make statistical forecasts of future performance levels
- Problem: *What sample?*
  - Independence
  - Uniform population

Sample Issues

- Independence
  - Network performance depends on the time at which it is measured
  - no independence
- Uniform population
  - Different transfer sizes constitute different populations for the purposes of prediction
  - *What transfer sizes should we use to measure network performance?*

The Network Weather Service



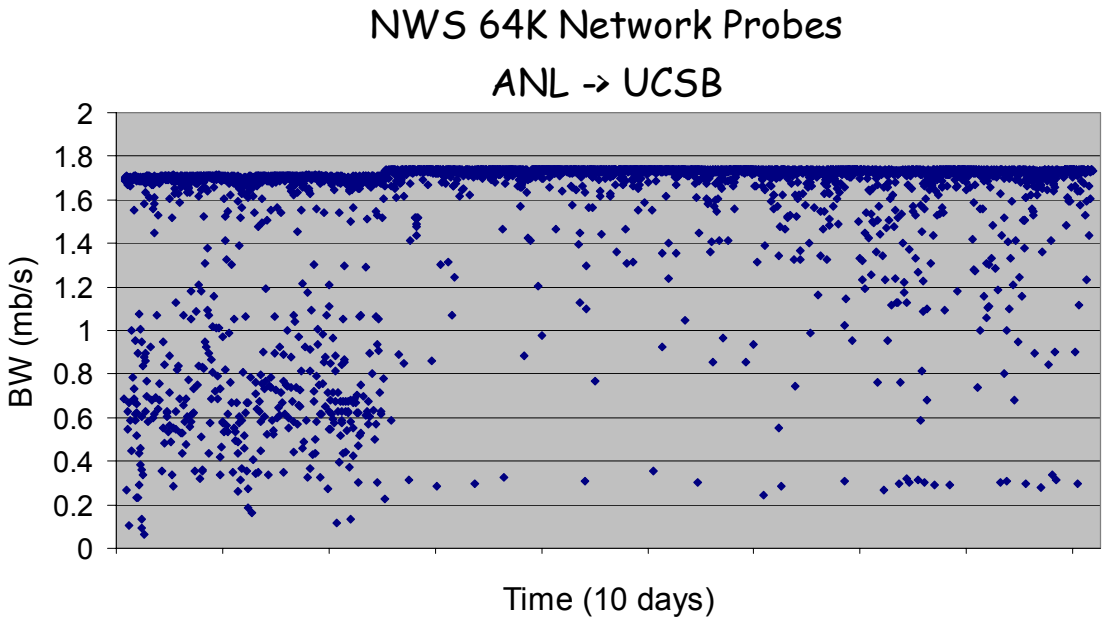
Our World: On-line Forecasting

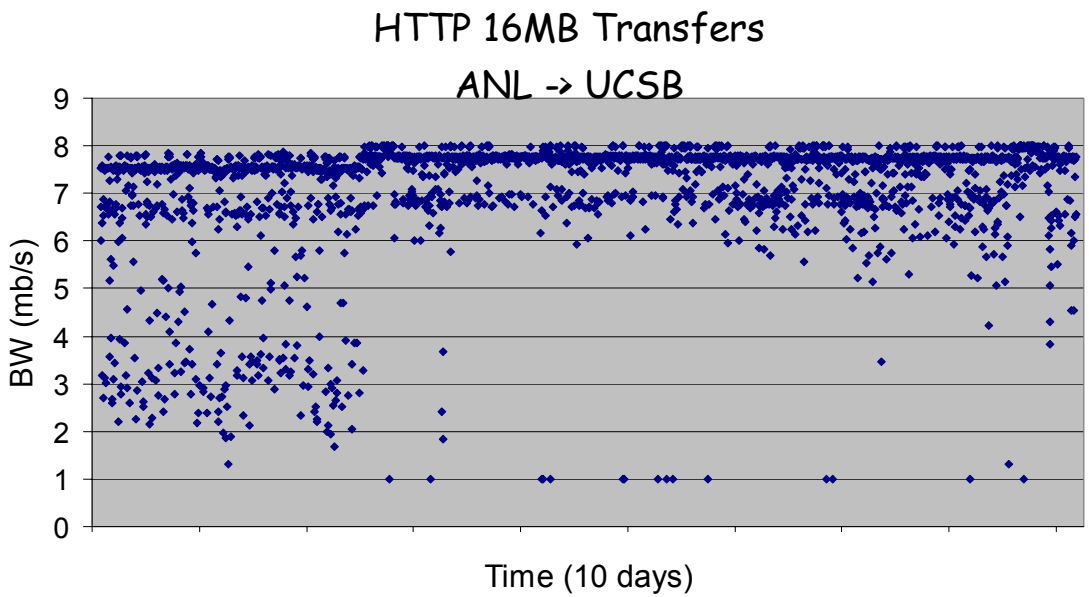
- Badly behaved autocorrelation doesn't mean "unpredictable."
  - Short-term forecasts are possible
- Our Approach:
  - Non-parametric or semi-non-parametric time series analysis using a constantly updated history
  - Conditional forecasting => fresh data implies a fresh forecast
- Univariate: predict only one transfer size per time series

Intrusiveness

- Can't probe the network with all possible transfer sizes
- Long transfers of one fixed size are even too much
- Question: *Can we use short, non-intrusive network probes to predict the future performance of long, intrusive network transfers?*

Network Rorschach Diagrams





Regression

- Two time series
  - Non-intrusive, frequent measurements (independent variable)
  - Intrusive, infrequent measurements (dependent variable)
  - For each intrusive measurement there is a "simultaneous" non-intrusive measurement
- Regression model: function that describes the dependency relationship

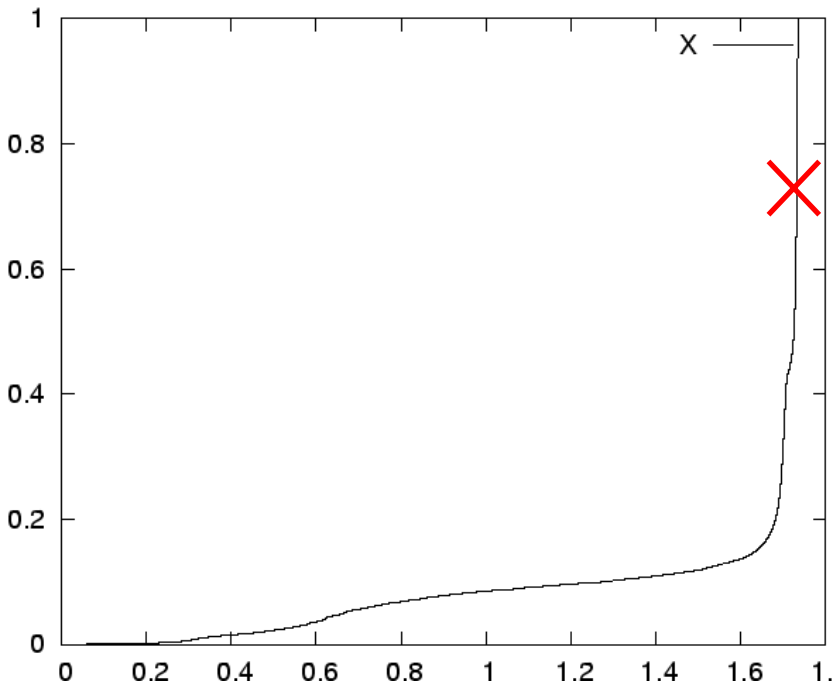
For Example

- Instrumentation data from GridFTP transfers yields a series of infrequent, long network transfers
- Periodic bandwidth probes (ala the Network Weather Service) yields a series of frequent, non-intrusive measurements
- Least-squares linear regression over matched transfers to calculate regression function
  - *S. Vazhkudai, J. Schopf, HPDC-11*
- Not very satisfying

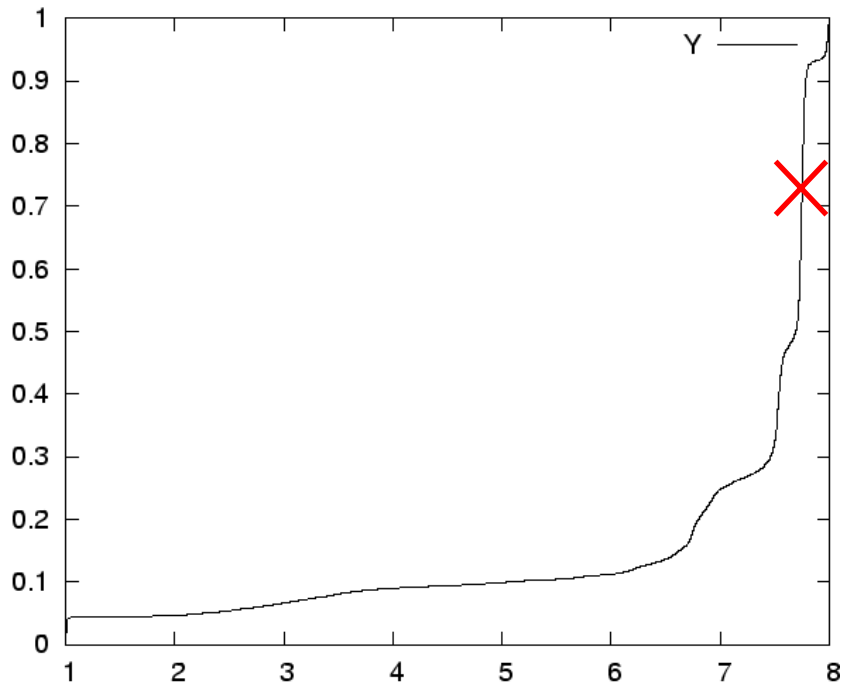
Another Approach

- Rank correlation: use the relative position of a measurement with respect to its observed population as a regression function
  - Assume that the quantiles are correlated
- For example: a non-intrusive short measurement that is bigger than 99% of all non-intrusive short measurements seen so far implies that the simultaneous long measurement will be bigger than 99% of all observed long measurements seen so far.

Easier to See with CDFs



64K NWS Transfers

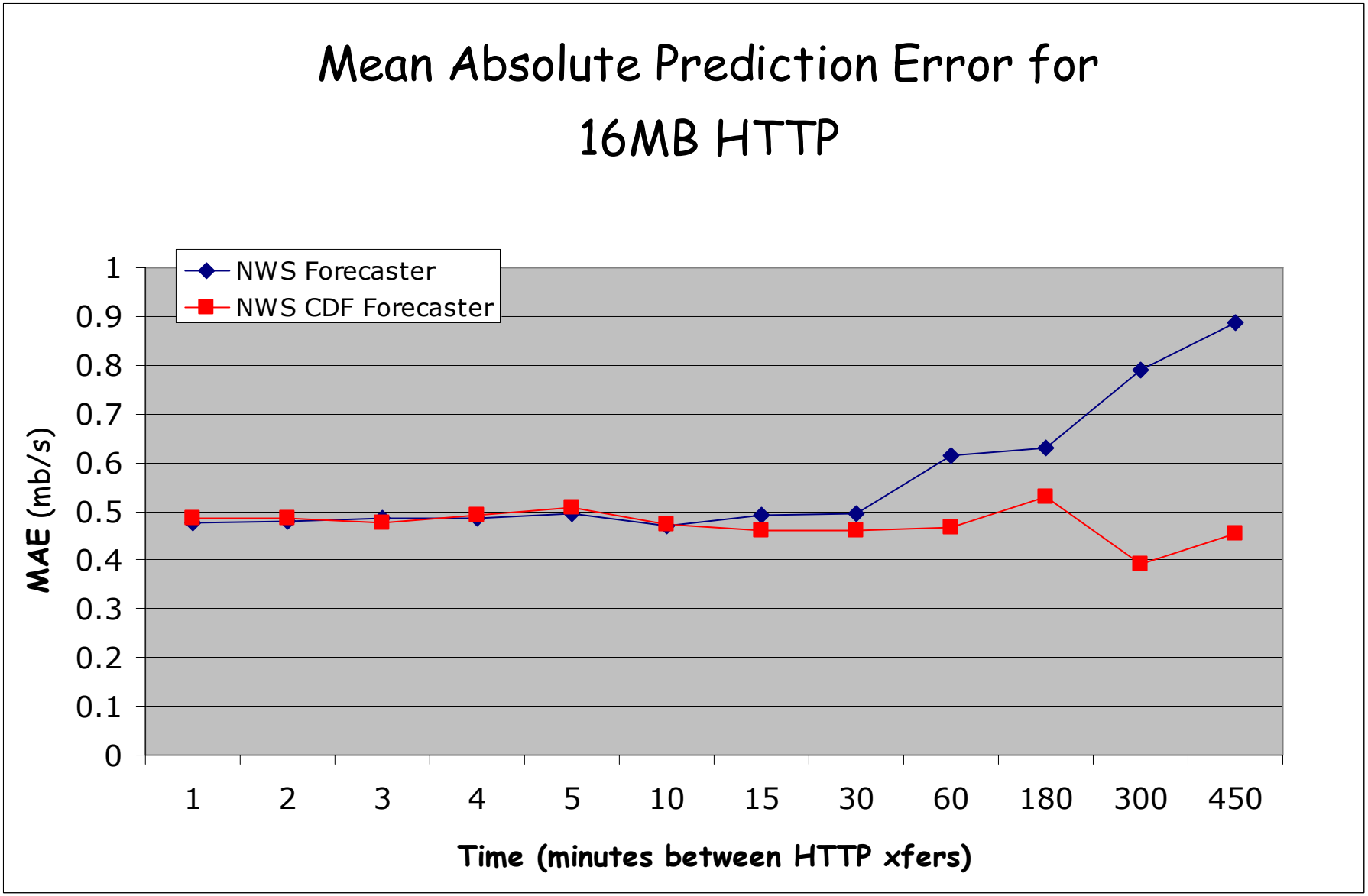


16M HTTP Transfers

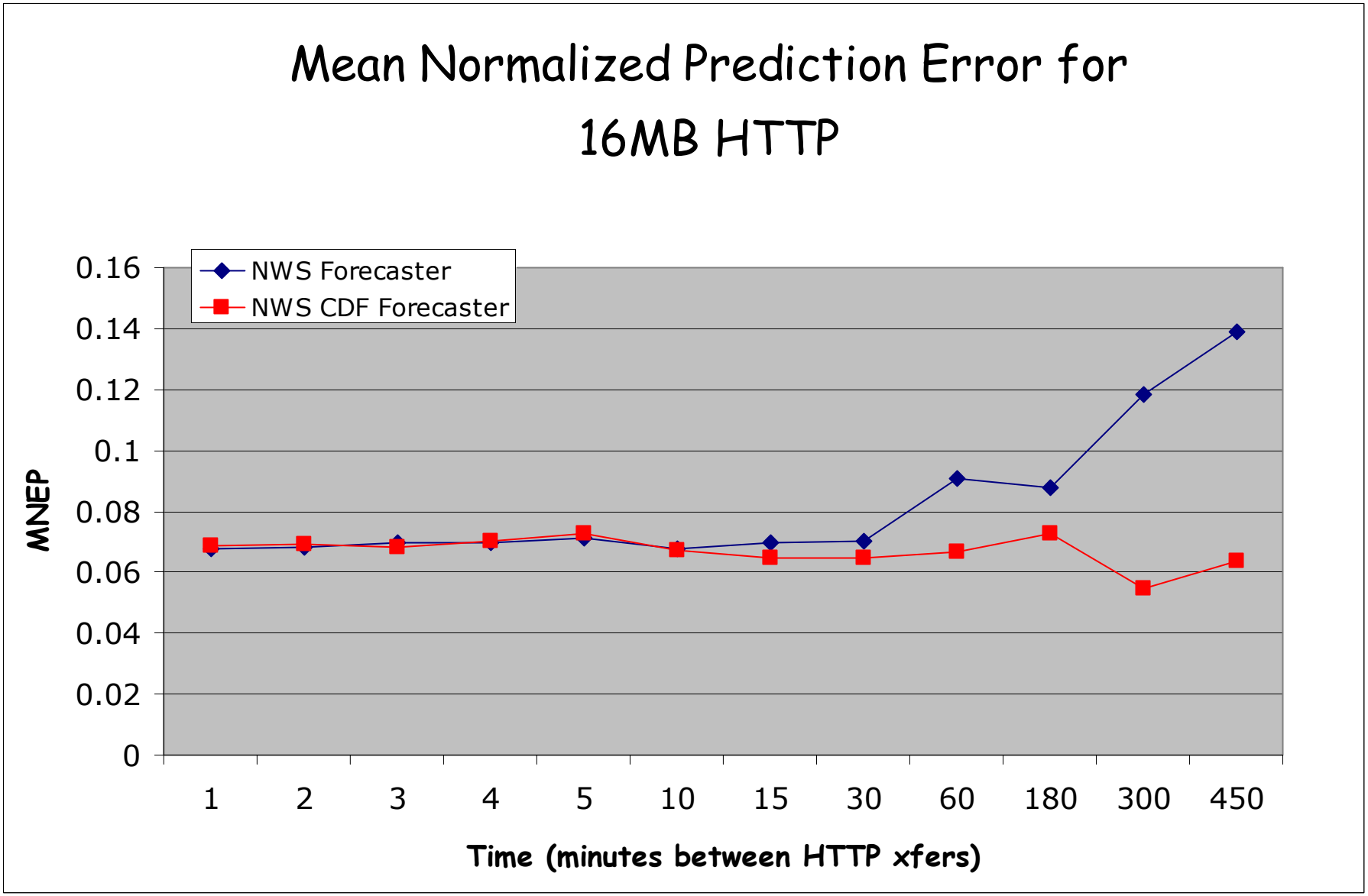
Measuring Forecast Accuracy

- Generate a forecast
- Compare the forecast to the measurement it predicts
- MAE: Mean Absolute Error
  - Average absolute difference
- MSE: Mean Square Error
  - Average of the square of the difference
- MNEP: Moving Normalized Error Percentage

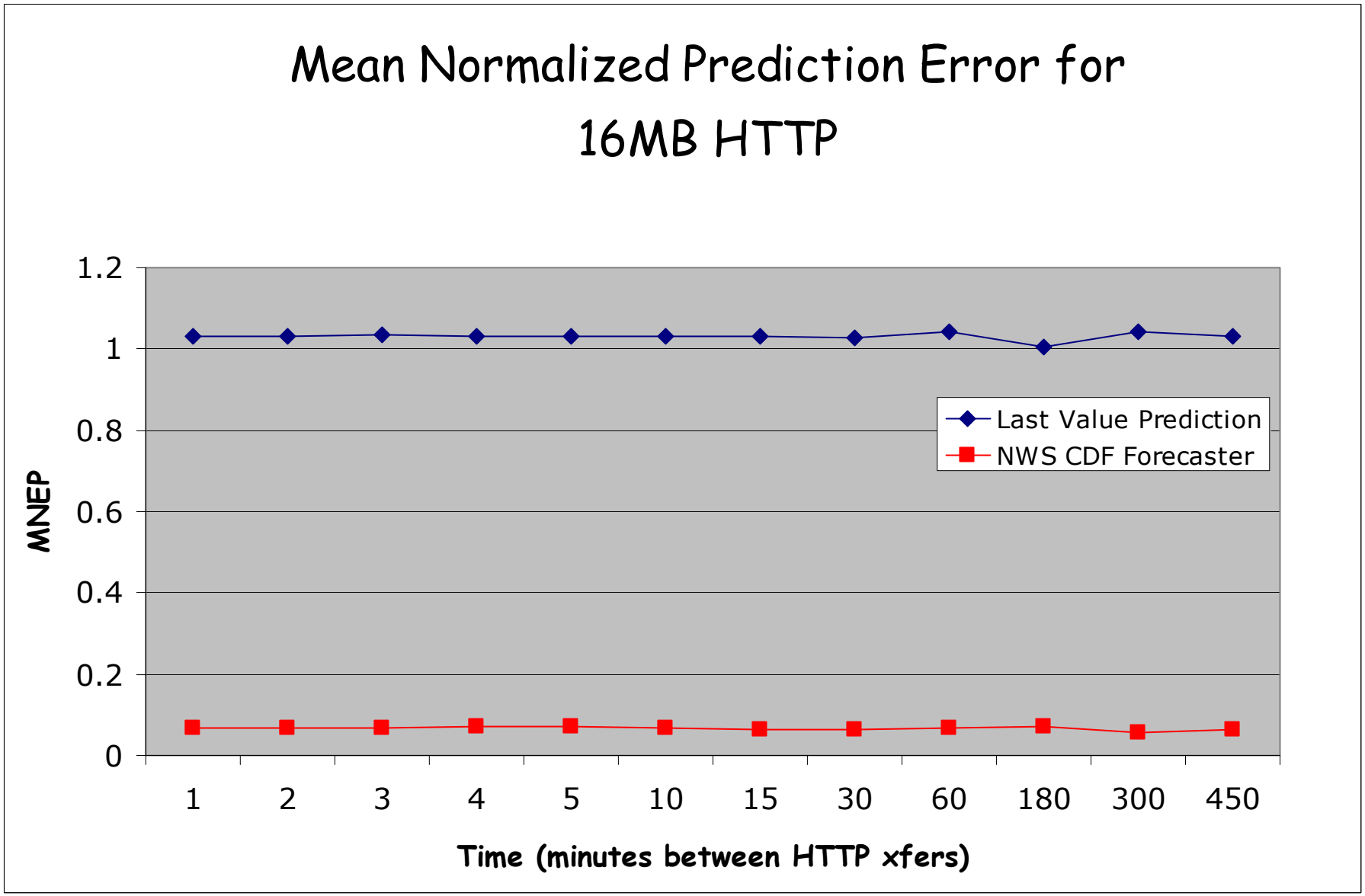
Old versus New Forecaster Accuracy



Error Percentage



What if you Used the Last Value?



What do Short BW Transfers Tell Us?

- Use the NWS forecast to determine rank in non-intrusive sample: NWS-Forecast-rank
- Find corresponding rank in intrusive sample
- Short transfers can generate forecasts out to 500 minutes with better than 10% accuracy (MNEP)

Credit and Thanks

- The **NWS** Project
  - staff and students
    Martin Swany, Graziano Obertelli, Matthew Allen, Wahid Chrabakh, Imran Patel, Vladimir Veytser
  - organizations
    SDSC, NCSA, The Globus Project (ISI/USC), The Legion Project (UVa), Condor project, MetaExchance Software Inc.
  - support
    NSF, NPACI, NASA, DARPA, USPTO, DOE

http://nws.cs.ucsb.edu